WU #2 - Wrangling

Math 154 - Fall 2017 - Jo Hardin

Solutions

Consider the diamond dataset. Below are 2 tasks which can be accomplished using the following syntax. **Identify the verbs and arguments for accomplishing each task** (x, y, z are length, width, and depth in mm). Note, you may not need the last `arrange`, but it won't cause errors. (From **Data Computing**, Daniel Kaplan)

```
diamonds %>%
  verb1( args1 ) %>%
  verb2( args2 ) %>%
  arrange( args3 ) %>% head(1)
```

```
require(ggplot2)
require(dplyr)
names(diamonds)
```

```
##  [1] "carat"   "cut"     "color"   "clarity" "depth"   "table"
##  [7] "price"   "x"       "y"       "z"
```

```
head(diamonds,2)
```

```
## # A tibble: 2 x 10
##   carat     cut color clarity depth table price     x     y
##   <dbl>   <ord> <ord>   <ord> <dbl> <dbl> <int> <dbl> <dbl>
## 1  0.23   Ideal     E     SI2  61.5    55   326  3.95  3.98
## 2  0.21 Premium     E     SI1  59.8    61   326  3.89  3.84
## # ... with 1 more variables: z <dbl>
```

1. Which color diamonds seem to be the largest on average (in terms of carats)? [I use the word "seem" because this is simply one dataset, and maybe it isn't representative of all diamonds. That is, the largest average color in this sample may not be the largest average color in the population.]

```
diamonds %>%
  group_by( color ) %>%
  summarize( avesize = mean(carat) ) %>%
  arrange( desc(avesize) ) %>% head(1)
```

```
## # A tibble: 1 x 2
##   color  avesize
##   <ord>    <dbl>
## 1     J 1.162137
```

2. What is the average price per carat of diamonds that cost more than $10,000?

```
diamonds %>%
  filter(price > 10000) %>%
  summarise( mean.ppc = mean(price/carat) ) %>%
  arrange( desc(mean.ppc) ) %>% head(1)
```

```
## # A tibble: 1 x 1
##   mean.ppc
##      <dbl>
## 1  8044.44
```